

# Content Validity Report for Rivet Education's Scoring and Evidence Guide (SEG)

Melanie Livet, Ph.D.  
Caryn S. Ward, Ph.D.  
Amelia Krynski

## Executive Summary

Rivet Education developed a Professional Learning Partner Guide (PLPG), a searchable database of learning providers with expertise in the adoption and implementation of High Quality Instructional Materials (HQIM), in response to the lack of resources and guidance available to support districts with evaluation of professional learning (PL) services. To be selected as a PL provider (PLP) in the PLPG, PL organizations have to submit an application that is reviewed and scored by Rivet Education using a Scoring and Evidence Guide (SEG).

To ensure confidence in the PLPG as a robust database of high quality PL services, the SEG needs to be evaluated as a reliable and valid assessment instrument. The National Implementation Research Network (NIRN) is conducting psychometric testing of the SEG, with this first phase focusing on establishing content validity.

Content validity is defined as the degree to which a tool or instrument is relevant to and representative of the concept(s) it is designed to measure. Data sources for assessing content validity included a survey and two focus groups, with a sample of PL providers, consumers of PL (i.e., state and districts), coaching researchers, and implementation science experts (N=24). Of interest were the frequency of the review cycles for submissions and renewals; the relevance, clarity, and alignment of the SEG gateways, constructs, and indicators (items), including any improvements that could be made; and the content, comprehensiveness, and utility of the SEG Scoring Rubric, including scoring criteria and weighting of indicators.

Overall, the results supported the content validity of the SEG, with high scores on the Content Validity Index (CVI) and Inter-rater Agreement (IRA) indices for clarity and relevance of gateways, constructs, and indicators (except for the Gateway 2 System Design and Leadership Support clarity CVI).

Alignment of indicators with gateways and constructs varied significantly, with no discernable patterns guiding re-alignment. However, edits were made to further clarify indicators rating below 50%. Furthermore, while comprehensiveness of the SEG Scoring Rubric was highly rated, participants disagreed on the weighting approach for the indicators.

A number of suggestions for improving the SEG emerged from the very detailed survey comments, the focus groups information, and a review of all data sources by the Rivet and NIRN teams, including recommendations on frequency of submissions and renewals, specific changes needed related to the gateways, constructs, and indicators, and a new approach to the Scoring Rubric.

In summary, the SEG was representative of the concepts it was designed to measure, including levels of content and HQIM expertise, quality of professional learning by type (i.e., adoption, launch, ongoing PL for teachers, and Systems Design and Leadership Support), and use of data for planning and improvement. Recommendations for refining the instrument were provided, with modifications being made by Rivet Education to produce Version 3.0 of the SEG. The next phase will be focused on establishing the psychometric properties of the SEG, by assessing inter-rater reliability and structural validity.

02

Introduction

04

Methods

06

Key Insights &  
Recommendations

11

Conclusion & Next  
Steps

# Introduction

Professional learning is recognized widely in the field of education as a critical implementation strategy to support teacher knowledge and skill development in instructional content, pedagogy, and use of instructional strategies and high quality instructional materials. Districts, on average, spend \$18,000 per teacher annually on synchronous professional learning (TNTP, 2015). However, the results are mixed when one examines the research based on the effectiveness of professional learning. For example, Blazar et al. (2019) found no difference in math textbook efficacy within schools across states with varied levels of professional development (dosage of professional learning ranging from 1.5 to 6 days). Furthermore, Boston Consulting Group (2014) reported the methods of professional development used by districts and schools are “highly fragmented” and often reported by teachers to be irrelevant and ineffective.

On the other hand, some studies yielded positive results related to the impact of professional development on the use of curricula. For example, professional development focused on curricula has been associated with above-average student gains in Science, Technology, Engineering, and Math instruction (Hill et al., 2020). Lynch et al. (2019) reported similar positive effects for professional learning focused on the use of curriculum materials in addition to focusing on improving content knowledge and pedagogy. Kaufman et al. (2020) reported that the perceived helpfulness of professional learning focused on curriculum use appeared to be more strongly linked to the use of standards-aligned practices than the frequency of professional learning.

The lack of consistent data on professional development effectiveness may be due in part to the still-emerging literature on its core components. A metasynthesis by Dunst et al. (2015) and a systematic review by Darling-Hammond et al. (2017) revealed common conditions or core features for professional development to be effective.

Core Features of Effective Professional Learning	
Dunst et al. (2015) metasynthesis: <ul style="list-style-type: none"><li>• trainer introduction, demonstration, and explaining of benefits of mastering content knowledge or practice</li><li>• active and authentic learning experiences</li><li>• opportunities to reflect on learning experiences</li><li>• coach or mentor support and feedback during the in-service training</li><li>• extended follow-up support to reinforce in-service learning</li><li>• in-service and follow-up supports of sufficient duration and intensity</li></ul>	Darling-Hammond et al. (2017) systematic review: <ul style="list-style-type: none"><li>• content-focused: professional learning focus on teachers associated with specific curriculum content and pedagogy</li><li>• incorporates active learning: engages teachers in designing and trying teaching strategies, using authentic artifacts, and interactive activities</li><li>• uses models of effective practices</li><li>• provides coaching and expert support</li><li>• offers feedback and reflection</li><li>• is of sustained duration – adequate time to learn, practice, use, and reflect upon to facilitate changes in learning</li></ul>

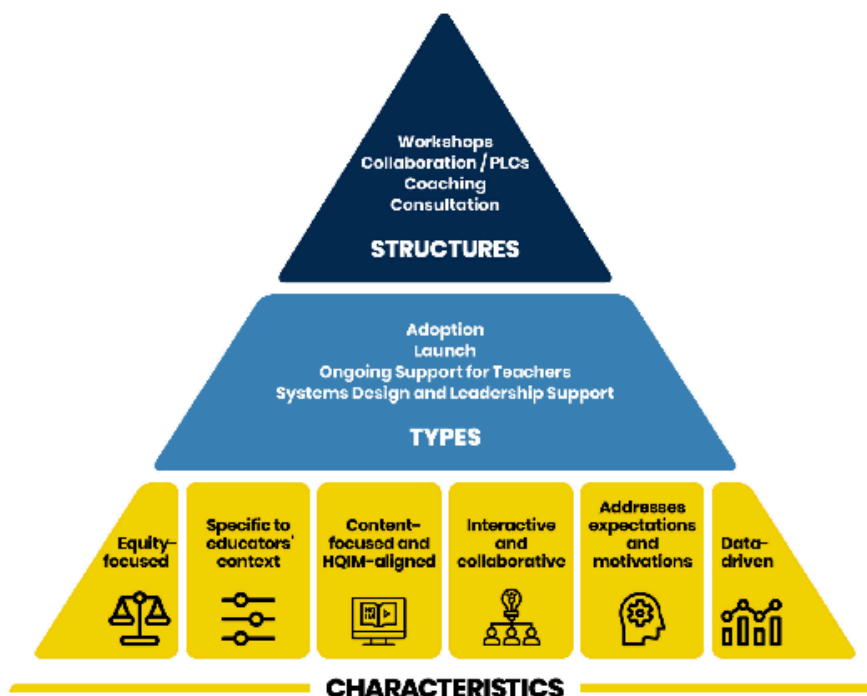
Based on the state of the research literature and given the costs associated with professional development, how does a district know their investment is likely to result in benefit? How does a district evaluate the wide range of services available and feel confident that the conditions or core features of effective professional learning are present? Few resources or guidance are available to support districts and schools in evaluating professional learning offerings (Student Achievement Partners, 2020).

Rivet Education recognized the gap in having reliable and accessible information to guide selection and evaluation of professional learning services. Specifically, Rivet developed the Professional Learning Partner Guide (PLPG), a national as well as state specific searchable database of providers who have expertise to support the adoption and implementation of HQIM. The goals of the PLPG are to 1) grow school district demand for high quality professional learning by building educators’ understanding of what it takes to deliver clear and effective instruction to teachers and pointing them to partners equipped to do so; and 2) improve the quality of the professional learning marketplace by signaling what constitutes high-quality, curriculum-based professional learning.

To support attainment of these goals, Rivet has developed a Framework for High Quality Professional Learning based on synthesis of research on high quality instruction, adult learning, educator professional learning, and implementation science (see Figure 1). The Framework then serves as the basis of Rivet’s review process and PLPG.

To be selected into the PLPG, PL providers have to complete a rigorous application. The application is reviewed using a Scoring and Evidence Guide (SEG) that was developed by Rivet for evaluation purposes. The reviews are conducted by Rivet’s trained team of teachers and leaders at school, district, and state levels, who have demonstrated expertise in HQIM and aligned professional learning. Currently, the review process using the SEG has identified 45 Professional Learning Providers for inclusion in the PLPG. The PLPG is in use within 5 states as of February 2022.

Figure 1. Framework for High Quality Professional Learning



Given the stability of the tool and review process, as well as its growth in use, the PLPG’s SEG is ready for a rigorous examination of its psychometric properties. The purpose of establishing reliability and validity of the SEG is to ensure confidence in its use as an evaluation guide for high quality professional learning services. In addition, psychometric data can be used to inform improvements. This assessment will result in a strengthened validated SEG and PLPG selection process.

This report presents a summary of data obtained during Phase I of the proposed psychometric evaluation, namely content validity testing. Content validity is defined as the degree to which a tool or instrument is relevant to and representative of the concept(s) it is designed to measure. In this context this study seeks to address the following question:

**Is the SEG an accurate reflection of the concepts it claims to measure, namely its three gateways (i.e., Content and HQIM expertise, Quality of Professional Learning by type–i.e., adoption, launch, ongoing PL for teachers, and Systems Design and Leadership Support–and using Data to Plan and Improve)? How can the tool be strengthened based on these data?**

The Phase I testing will result in a new version of the SEG (3.0), which will then be tested in Phase 2.

# Methods

---

A mixed methods approach was adopted to establish the content validity of the SEG Version 2. Data sources included: (1) a 4-section survey that consisted of quantitative and qualitative items, administered to 24 participants, including PL providers, consumers of PL, PL and coaching researchers, and implementation science experts; and (2) two focus groups, with 5-6 participants each, that included PL providers, consumers of PL (ie, state and districts), and implementation science experts. This study, including the study design, development of data collection tools and surveys, participant recruitment, survey administration, focus group facilitation, analysis and data pulls, and summary preparation, was conducted between November and April 2024. Upon completion of the analysis and preparation of summary documents, the findings were discussed in detail with the Rivet team during a 2-day in-person meeting and one 2-hour virtual meeting in April 2024.



**Participant Recruitment and Sample Description:** Survey participants from all five target audiences were recruited by the Rivet and NIRN teams based on their access to experts and their extensive networks of users. Recruitment continued until a sample of 30 was attained (15 PL providers, 10 PL consumers, and 5 researchers/experts). Focus group participants self-selected following the completion of the survey. Participants were provided with \$100 incentives upon survey completion, and \$100 for participating in the focus groups.

A total of 23 participants completed the survey, with 21 finishing all four sections. Based on demographic information, the majority were female (83%, N=23), White (83%), with a Master's degree or above (96%), and more than 20 years experience working in education (74%). Forty-eight percent self-identified as PL providers, 17% as district administrators, 17% as state Department of Education (DOE) staff, and 17% as researchers. Of the 11 participants in the focus groups, 82% were female, 100% were White, 100% had a Master's degree or above, and 73% had 20 or more years of experience working in education. Fifty-five percent self-identified as PL providers, 9% as district administrators, 27% as state DOE staff, and 9% as researchers.

---

**Survey Administration:** The survey was designed as an online instrument to assess the SEG's content validity. More specifically, data were meant to evaluate and inform any needed changes to: (1) the administration of the SEG based on frequency and timing; (2) SEG gateways and constructs, based on relevance and clarity ratings; (3) the SEG items or indicators, based on relevance, clarity, and alignment with the gateways and/or constructs; and (4) the scoring rubric, based on comprehensiveness and weighing ratings. Relevance (i.e., representativeness) and clarity ratings for gateways, constructs, and indicators, as well as comprehensiveness of the scoring criteria for each indicator, were assessed on a 4-point scale, with 1 being "not at all [relevant/clear/comprehensive]", 2 "requiring major revisions," 3 "requiring minor revisions," and 4 "leave as is." Alignment of the indicators with the corresponding gateway/construct was accomplished by having participants match each indicator to the list of gateway/constructs. Finally, weighting for each indicator was assessed by having participants select whether a particular indicator had to be weighted less, more, or similarly to other indicators in that section. Participants also indicated the types of changes, deletions, and additions needed through open-ended comments. The survey was administered in January and February, 2024, with participants being given approximately two to three weeks to complete.

---



**Focus Groups Administration:** Focus group participants were assigned to one of two focus groups based on availability and role. The focus groups were facilitated via zoom by a trained member of the NIRN team using a semi-structured protocol designed to elicit additional insights into the content, utility, and use of the guide. Two of the team members were taking notes, with the focus groups also being transcribed for analysis. The focus groups were conducted in March, 2024.

---

**Quantitative and Qualitative Analysis:** A data plan was prepared prior to survey development to guide the analysis of all data sources, and later used to organize information once analyzed. Both Interrater Agreement (IRA) and the Content Validity Index (CVI) were used to analyze the relevance, clarity, and comprehensiveness survey data (Rubio et al., 2003). The IRA is an indicator of the extent to which participants are reliable in their ratings. Interrater agreement was assessed for each “criterion” (gateways, constructs, indicators) by dividing the responses into dichotomous ratings (one or two vs. three or four). The number of ratings indicating agreement was counted and divided by the total number of raters to calculate IRA, with a goal of  $\geq 80\%$  agreement. The CVI represents the extent to which the “criterion” measured is adequate in terms of relevance, clarity, and comprehensiveness across all respondents. It was calculated as a mean score on a scale of 1 to 4, with higher scores indicating fewer or no revisions (with a goal of  $> 3.0$ ). Alignment of each indicator to the relevant gateway/construct was computed as the percent of positively matched responses over the total number of matched and non-matched responses for that indicator, with the goal being above 50%. A score of 50% or less (indicating misalignment) triggered a review of that indicator. For ease of discussion and review with the Rivet team, open-ended comments were grouped by the specific questions that they were addressing (e.g., Gateway 1, Construct 3).

Focus group data were analyzed using thematic analysis. A codebook with a-priori codes based on the questions (e.g., what they liked, what needed improvement) was created for application during a review of the transcripts, with final analysis including sub-themes within each code (e.g., administration process, overall conceptual structure). This information was summarized into a word document that organizes the findings by theme. The analysis was conducted by a member of the NIRN team, with another member reviewing for accuracy. Information from the focus groups was combined with information from the survey, including the open ended comments, and reviewed in detail during the meetings with the Rivet team.

---



# Key Insights & Recommendations

## SEG Administration

### Recommendations

- ▶ Maintain biennial renewal cycles (every 2 years) pending no other changes for those PLPs already in the PLPG
- ▶ Add a grandfathering clause so that applicants do not have to re-apply as a result of external changes impacting the scoring, such as SEG modifications or changes in educational standards
- ▶ Consider abbreviated applications for renewal submissions, targeting gaps and low scored domains
- ▶ Publicize available screener (that Rivet just added to its newer application cycle) so that applicants can determine their level of readiness for submitting an application
- ▶ Consider adding instructions and tips for new applicants and organizations with fewer resources into existing guidance webinars and materials

### Satisfaction with Overall Process

- While focus group participants highlighted the need for a standardized process to evaluate Professional Learning Providers (PLPs), they also pointed out that the application process was time- and resource-intensive, with concerns about the return on investment (i.e., being listed in the PLPG). Requests were made to streamline the application process and SEG as much as feasible.
  - One participant recommended the use of a screener that would help applicants determine their level of readiness for submitting an application.
- Concerns were also raised among focus group respondents around the inherent bias associated with the application process, namely towards larger, more established PLPs with resources necessary to complete the application.

### Frequency of Reviews for Application Revision and Resubmission

- Survey data indicates ranges from 30 days to one year to variable depending on the revisions required for an application to earn a passing score.

### Frequency of Renewal Reviews

- Based on survey responses (N=23), the majority of respondents (60.9%) advocated for an every 2 year cycle of reviews of qualifications and services provided by the PLPs that are listed in the PLPG.
- Based on the survey suggestions for "other" (30.4%), responses ranged from every 3 years, to based on PLP changes in services, to a graduated schedule (more frequently at the beginning).
- Some focus group participants recommended a 5 year renewal process pending no significant changes to the PLP's offerings, content, or delivery and no changes external to the PLP that would impact scoring (e.g., changes in educational standards, changes in SEG criteria). Should these changes occur, it was suggested that Rivet Education consider a rolling renewal option.
- Others proposed an annual renewal check-in due to implementation drift and changes in PL provider organizations (e.g., turnover).
- All agreed that renewals should be abbreviated and focused, with only new materials being asked to be provided; for Gateway 3 of Using Data to Plan and Improve, it was also suggested that, depending on the PLP maturity level, the focus should be different for renewals. Some may be able to showcase efficacy and effectiveness of PL, while others may not be.





# Key Insights & Recommendations

## Gateways, Constructs, and Indicators: Relevance, Clarity, and Alignment

### Recommendations

- ▶ Revise the Gateway 1 screening form (used by applicants to submit information for Gateway 1) to collect information on knowledge of content pedagogy and how professional learning materials are customized for curricula published across multiple platforms
- ▶ Clarify instructions on Gateway 1 to ensure that applicants understand they need to pass either 1 of the 2 indicators based on the type of PL they are applying for... *unless* they are applying for multiple types of PL
- ▶ Further clarify the different constructs or PL types within Gateway 2 (e.g., adoption, launch, on-going professional learning) to reflect the implementation process (e.g., edit wording, reclassify certain indicators)
- ▶ Provide greater clarity on target audience and level of the system targeted for the gateways and relevant constructs and indicators
- ▶ Ensure student perspective is centered and captured within relevant indicators
- ▶ Create a glossary to provide definitions for key terms
- ▶ Revise indicators as needed to be stated in a positive manner
- ▶ Revise specific indicators as needed to provide greater clarity and specificity and reduce complexity
- ▶ Further define the purpose of Gateway 3 and distinction from Gateway 2
- ▶ Consider additional indicators to address evaluating coaching, sharing of data between Providers and Districts, use of data for monitoring and improvement, and communication and visible promotion of HQIM implementation

### Overall Conceptual Structure: Gateways, Constructs, Indicators

- Focus group participants were curious to learn more about the approach to adult learning grounding the SEG. Some mentioned that it did not necessarily align with the PLP vision for adult learning (e.g., strategies for assessing methods to address teacher biases through PL should play a more prominent role in the SEG).
- Others reported that while they liked the overall structure of the SEG, the process necessary to familiarize oneself with its structure and requirements to successfully apply was time intensive.

### Relevance and Clarity of Gateways

- The CVI, representing the extent to which the gateways were adequate in terms of relevance to the framework for High Quality Professional Learning and clarity across all respondents (N=23), was high (M=3.7 for relevance, and 3.5 for clarity, on a scale of 1 to 4, with higher scores indicating few or no revisions) (see Table 1).
- The IRA, assessing interrater agreement among all respondents (N=23) on scores for relevance and clarity of the Gateways, was also high, with 97.1% for relevance and 94.2% for clarity (see Table 1).
- The majority of survey participants agreed with the order sequence of the gateways with no suggested changes (N=21/22).

- Based on the qualitative comments on the gateways overall, the following was noted: (1) confusion as to the need for three gateways, connections between the gateways, and the overlap between types of PL in Gateway 2; (2) the need to account for levels of the system (i.e., target audience) as part of the gateways, with added specificity and definitions.
  - Gateway 1 comments highlighted the need for communicating to PLPG applicants as well as PL consumers that pedagogical knowledge is addressed through the questions being asked in Gateway 1. Adding a question explicitly about pedagogy knowledge was recommended.
  - Gateway 2 comments referenced the overlap in types of professional learning constructs being measured (i.e., adoption, launch, ongoing support, and system design and leadership support) contributing to a burdensome application process. Additional comments included operationalizing less defined yet commonly used terms such as “beliefs” and “all.”
  - Gateway 3 comments pointed to the need to 1) clarify the “data” needed to make improvements and be inclusive of different types of data; 2) the need for establishing a baseline for data quality with substantiated impact; 3) the need to broaden beyond the process of using data to including evidence of data use as well as efficacy and effectiveness; and 4) add sustainability planning with consumers of PL.

#### **Relevance and Clarity of Constructs (Appropriate for Gateway 2 only)**

- The CVI, representing the extent to which the constructs in Gateway 2 were adequate in terms of relevance to the gateway and clarity across all respondents (N=23), was high (ranging from M=3.5 to 3.8 for relevance, and 3.7 to 3.8 for clarity, on a scale of 1 to 4, with higher scores indicating few or no revisions) (see Table 2).
- The IRA, assessing interrater agreement among all respondents (N=23) on scores for relevance and clarity of constructs in Gateway 2, was also high, with ranges of 87% to 100% for relevance and 91.3% to 100% for clarity (see Table 2).
- Based on the qualitative comments on the constructs within Gateway 2, the following was noted:
  - Clarity is needed in the vocabulary and language used in the definition of the Adoption construct.
  - Greater distinction is needed between the focus of the professional learning within Launch and Ongoing Support, and inclusion of leadership within Launch, versus the role of leadership in the System Design and Leadership construct.
  - Greater emphasis is needed on building collective agreement or shared understanding of the need for HQIM.
  - Inclusion of integrity within all constructs in Gateway 2 and use of integrity versus fidelity is needed.
  - Greater clarity is warranted as to the specific target audience of leadership and level within the System Design and Leadership support construct.



## Relevance, Clarity and Alignment of Indicators

- The CVI, representing the extent to which the indicators within the gateways were adequate in terms of relevance to the construct and/or gateway and clarity across all respondents (N=23), was high (ranging from M=3.5 to 4.0 for relevance, and 2.9 to 3.9 for clarity, on a scale of 1 to 4, with higher scores indicating few or no revisions) (see Table 2). Several indicators within Gateway 2 for System Design and Leadership support fell below 3.0 with an overall average of 2.9 for clarity, with five of eight indicators ranging from 2.8 to 2.9.
- The IRA, assessing interrater agreement among all respondents (N=23) on scores for relevance and clarity of indicators, was also high, with ranges of 88% to 100% for relevance and 86% to 100% for clarity (see Table 2).
- Overall, several areas for improvement were identified for the indicators based on the qualitative comments. These included: 1) greater clarity on the term “equity” when used within indicators, 2) need to ensure indicators are streamlined and not referencing multiple concepts (e.g., double-barreled), and 3) greater specificity in language.
- Based on the qualitative comments about the indicators for each gateway, the following was noted:
  - Comments about the Gateway 1 indicators referenced the need to define “design principles” and the need for greater clarity on what was meant by “shifts” in content standards.
  - Gateway 2 overarching indicators comments referenced the need for indicators to be stated in a positive rather than deficit-based manner. There also was a lack of clarity on using student work to demonstrate Gateway 2 proficiency, especially for Adoption and Launch.
  - Gateway 2 indicators comments referenced the need for clarity on the phrases “in a variety of formats” and “learning progressions” in Adoption indicators and “arc of learning” in Launch to provide more specificity. In addition, edits were made to one of the Adoption indicators referencing “defining and refining a high quality instructional math vision” (changing it to “defining or refining”).
  - Gateway 2 indicators comments for Ongoing Professional Learning noted the need for greater focus on educators’ specific actions demonstrating a commitment to equity.
  - Gateway 2 System Design and Leadership Support indicators comments focused on the need to clearly define the target audiences within the indicators (e.g., district level leaders, school leaders). They also called for greater specificity on the information/data needed to support effective implementation and decisions for scaling and allocating additional resources.
  - Gateway 3 indicators comments highlighted the need for greater clarity on “evaluating impact of its services”; greater distinction between two particular indicators (3.3 and 3.4); greater specificity in the type of data/information to be used in planning improvements; and clarity on the use of “differentiation” within an indicator.
- The level of alignment of indicators with their constructs/gateways ranged from 13.6% to 100%, with the average across indicators being 66.7% (Table 3). A number of edits were made to improve clarity to those indicators with 50% or less alignment agreement. The majority of those edits were made to the Gateway 3 indicators to ensure greater specificity and improve alignment with the specified construct (e.g., the third indicator in Gateway 2-Launch was modified to specify the purpose of needing evidence of student learning). It should be noted that no indicators were changed with regards to the constructs they aligned with.



# Key Insights & Recommendations

## Scoring Rubric: Comprehensiveness/Weighting and Evidence

### Recommendations

- ▶ Consider weighting all indicators equally using a consistent 3-point scale, with a clear cutoff score representing a passing “grade” (e.g., 70%)
- ▶ Given equal weighting, retain only “core” indicators
- ▶ Further define the 3-point scoring system to include quality as well as presence/absence of a set of criteria, with: (1) maximum score indicates that all criteria have been met with high quality; (2) middle score indicates that criteria are fully met but with mediocre quality, OR criteria are partially met but with high quality; and (3) the lowest scores indicates that criteria are partially met with mediocre quality
- ▶ Consider expanding options for uploading diverse types of materials and providing additional guidance on examples of evidence that can be used for “hard-to-demonstrate” criteria such as soft skills (e.g., quality of relationships) and pervasive concepts (e.g., equity)

#### Comprehensiveness and Weighing

- Comprehensiveness of the scoring criteria for each indicator, as calculated by the CVI, ranged from 3.4 to 3.9 (N=21) (on a scale of 1 to 4, with higher scores indicating few or no revisions) (see Table 4). The Comprehensiveness IRA, assessing interrater agreement among all respondents (N=21), ranged from 90.5% to 100% (see Table 4).
- Weighting of each indicator was assessed by asking survey respondents whether the indicator was appropriately weighted, should be weighted less, or should be weighted more. The results were highly varied with no discernible patterns (see Table 5). It is worth noting that only 10 of the 37 indicators had an acceptable IRA (above 80%), indicating lack of consensus among respondents as to the weighting of the items.
- Based on survey comments and the focus group results:
  - There was confusion as to the rationale for the differences in rating scales across indicators (0, 1, 2 or 0, 2, 4).
  - Suggestions were made to consider quality of evidence in addition to presence or absence of scoring criteria.
  - Suggestions were made for scoring criteria edits to ensure positive language is used throughout and other criteria-specific edits.

#### Supporting Evidence (Types, Use)

- A number of comments and suggestions were made related to the types and formats of the evidence used in support of the scoring criteria to demonstrate proficiency, including:
  - the need to accommodate various formats of uploaded materials (e.g., videos, links).
  - the challenges associated with the specificity of scoring criteria, which is oftentimes demonstrated through aspects of various materials rather than one piece of evidence (e.g., equity).
  - the challenges associated with opportunities to showcase intangible aspects of the work, such as relationship building between providers and districts.
  - the need to account for the gap between the materials included in the application and what PL providers are actually doing in the field (e.g., “PL in action”), including personalization of PL offerings.
  - suggestions to expand the criteria to incorporate evidence related to PL users (districts) and beneficiaries (students, families).

## Conclusion and Next Steps

In summary, results supported the content validity of the SEG. Data from both the survey and the focus groups confirmed that the SEG was representative of the concepts it was designed to measure, including levels of content and HQIM expertise, quality of professional learning by type (i.e., adoption, launch, ongoing PL for teachers, and Systems Design and Leadership Support), and use of data for planning and improvement. Recommendations for improvements were identified to strengthen the frequency of the review process, the relevance, clarity, and alignment of the gateways, constructs, and indicators, and the accuracy of the scoring rubric. Of note, while this report summarizes high-level suggestions for improvement, the NIRN and Rivet teams reviewed the data from all sources in painstaking details over the course of a 2-day in-person meeting as well as follow-up virtual meetings with specific changes recommended for all aspects of the SEG (including wording of indicators). As a result, the SEG is being revised into version 3.0.

- ▶ **The NIRN team will continue to support the Rivet team in their use of the data to produce version 3.0 of the SEG.**
- ▶ **The NIRN team will facilitate the co-creation of a plan for the next phase of the evaluation, which is designed to establish the inter-rater reliability and structural validity of the SEG.**
- ▶ **The NIRN and Rivet teams will co-create the dissemination plan and communication campaign for sharing Phase I results with PL Providers, PL consumers, and researchers, which will include:**

**The development and publishing of a blog**

**The development and publishing of a brief**

**The development and hosting of a webinar to be offered during Rivet Education's PL Power hour**

## References

- Blazar, D., Heller, B., Kane, T., Polikoff, M., Staiger, D., Carrell, S.,...& Kurlaender, M. (2019). Learning by the book: Comparing math achievement growth by textbook in six Common Core states [Research Report]. Cambridge, MA: Center for Education Policy Research, Harvard University. [https://cepr.harvard.edu/files/cepr/files/cepr-curriculum-report\\_learning-by-the-book.pdf](https://cepr.harvard.edu/files/cepr/files/cepr-curriculum-report_learning-by-the-book.pdf)
- Boston Consulting Group. (2014). Teachers know best: Teachers' views on professional development. The Bill & Melinda Gates Foundation. <https://usprogram.gatesfoundation.org/-/media/dataimport/resources/pdf/2016/11/gates-pdmarketresearch-dec5.pdf?rev=f770d1b3de574f8c855c1428811ba30f&hash=991179D78E52CC037B0D0B4CA00EE9EE>
- Darling-Hammond, L., Hyster, M. E., Gardner, M., & Espinoza, D. (2017). Effective teacher professional development. Learning Policy Institute. <https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report>
- Dunst, Carl J., Bruder, Mary Beth, & Hamby, Deborah W. (2015). Metasynthesis of in-service professional development research: Features associated with positive educator and student outcomes. *Educational Research and Reviews*, 10(12), 1731-1744. doi:10.5897/ERR2015.2306
- Hill, H. C., Lynch, K., Gonzalez, K. E., & Pollard, C. (2020). Professional development that improves STEM outcomes. *Phi Delta Kappan*, 101(5), 50-56. <https://doi.org/10.3102/0162373719849044>
- Kaufman, J. H., Doan, S., Tuma, A. P., Woo, A., Henry, D., & Lawrence., R. A. (2020). How instructional materials are used and supported in U.S. K-12 Classrooms. Rand Corporation. <https://doi.org/10.7249/RRA134-1>
- Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, 41(3), 260–293. <https://doi.org/10.3102/0162373719849044>
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94-104. <https://doi.org/10.1093/swr/27.2.94>
- Student Achievement Partners. (2020). Principles for high-quality, standards-aligned professional learning. Achieve the Core. <https://achievethecore.org/content/upload/Principles%20for%20High-Quality,%20Standards-Aligned,%20Professional%20Learning.pdf>
- TNTP. (2015). The mirage: Confronting the hard truth about our quest for teacher development. Brooklyn, NY: Author. <https://tntp.org/publication/the-mirage-confronting-the-truth-about-our-quest-for-teacher-development/>

## Appendices

**Table 1:** CVI and IRA for Gateways

	Relevance (to Framework)		Clarity	
	CVI (mean)	IRA (%)	CVI (mean)	IRA (%)
<b>Gateways</b>				
Gateway 1	3.7	95.7%	3.5	95.7%
Gateway 2	3.7	100%	3.5	95.7%
Gateway 3	3.7	95.7%	3.6	91.3%
<b>Framework Gateways Average</b>	3.7	97.1%	3.5	94.2%

**Table 2:** CVI and IRA for Constructs and Indicators

	Relevance		Clarity	
	CVI (mean)	IRA (%)	CVI (mean)	IRA (%)
<b>CONSTRUCTS (Gateway 2)</b>				
Adoption	3.5	87%	3.7	91.3%
Launch	3.8	95.7%	3.7	100%
Ongoing Professional Learning for Teachers	3.8	100%	3.8	100%
System Design and Leadership Support	3.7	91.3%	3.8	95.7%
<b>Gateway 2 Construct Average</b>	3.7	93.5%	3.8	96.8%
<b>Indicators</b>				
Indicator 1.1	3.9	95.7%	3.7	95.7%
Indicator 1.2	3.7	92.3%	3.6	91.3%
<b>Gateway 1 Indicator Average</b>	3.8	94%	3.7	93.5%

Indicator 2A.1	3.8	95.7%	3.9	100%
Indicator 2A.2	3.8	100%	3.7	95.7%
Indicator 2A.3	3.6	87%	3.7	91.3%
Indicator 2A.4	3.9	100%	3.7	95.7%
Indicator 2A.5	3.8	95.7%	3.6	91.3%
Indicator 2A.6	3.7	91.3%	3.7	91.3%
<b>Gateway 2 Adoption Indicator Average</b>	3.8	95%	3.7	94.2%
Indicator 2L.1	4	100%	3.9	95.7%
Indicator 2L.2	3.6	91.3%	3.3	87%
Indicator 2L.3	3.7	87%	3.7	91.3%
Indicator 2L.4	4	100%	3.7	95.7%
Indicator 2L.5	3.7	95.7%	3.7	91.3%
Indicator 2L.6	3.9	100%	3.6	95.7%
Indicator 2L.7	3.9	100%	3.7	95.7%
Indicator 2L.8	3.8	95.7%	3.7	91.3%
Indicator 2L.9	3.8	95.7%	3.7	95.7%
<b>Gateway 2 Launch Indicator Average</b>	3.8	96.2%	3.7	93.3%
Indicator 2T.1	4	100%	3.8	95.7%
Indicator 2T.2	3.6	91.3%	3.4	87%
Indicator 2T.3	3.9	95.7%	3.7	91.3%
Indicator 2T.4	4	100%	3.7	95.7%
Indicator 2T.5	3.9	100%	3.7	95.7%
Indicator 2T.6	4	100%	3.7	95.7%
Indicator 2T.7	3.9	100%	3.7	95.7%
<b>Gateway 2 Ongoing Professional Learning for Teachers Indicator Average</b>	3.9	98.1%	3.7	93.8%
Indicator 2S.1	3.9	100%	3	95.5%



Indicator 2S.2	3.6	95.5%	2.8	90.9%
Indicator 2S.3	3.6	86.4%	2.9	90.9%
Indicator 2S.4	4	100%	2.9	95.5%
Indicator 2S.5	3.9	100%	2.9	95.5%
Indicator 2S.6	3.9	100%	2.9	90.9%
Indicator 2S.7	4	100%	3	95.5%
Indicator 2S.8	3.9	95.5%	3	95.5%
<b>Gateway 2 System Design and Leadership Support Indicator Average</b>	3.9	97.2%	2.9	93.8%
<b>Gateway 2 Indicator Average (across all Gateway 2 constructs)</b>	3.9	96.6%	3.5	93.8%
Indicator 3.1	3.9	100%	3.8	95.7%
Indicator 3.2	3.8	100%	3.7	100%
Indicator 3.3	3.8	95.7%	3.8	100%
Indicator 3.4	3.9	95.7%	3.7	95.7%
Indicator 3.5	3.6	91.3%	3.5	91.3%
<b>Gateway 3 Indicator Average</b>	3.8	96.5%	3.7	96.5%

**Table 3:** Indicators - Alignment to relevant Gateways/Constructs

	Alignment (%)
<b>Gateway 1 Indicators:</b>	
Indicator 1.1	68.2%
Indicator 1.2	45.5%
<b>Gateway 2 Adoption Indicators:</b>	
Indicator 2A.1	81.8%
Indicator 2A.2	86.4%
Indicator 2A.3	13.6%
Indicator 2A.4	90.9%

Indicator 2A.5	31.8%
Indicator 2A.6	68.2%
<b>Gateway 2 Launch Indicators:</b>	
Indicator 2L.1	77.3%
Indicator 2L.2	100%
Indicator 2L.3	40.9%
Indicator 2L.4	95.5%
Indicator 2L.5	27.3%
Indicator 2L.6	36.4%
Indicator 2L.7	40.9%
Indicator 2L.8	40.9%
Indicator 2L.9	45.5%
<b>Gateway 2 Ongoing Professional Learning for Teachers Indicators:</b>	
Indicator 2T.1	90.9%
Indicator 2T.2	95.5%
Indicator 2T.3	45.5%
Indicator 2T.4	95.5%
Indicator 2T.5	77.3%
Indicator 2T.6	54.5%
Indicator 2T.7	63.6%
<b>Gateway 2 System Design and Leadership Support Indicators:</b>	
Indicator 2S.1	81.8%
Indicator 2S.2	90.9%
Indicator 2S.3	54.5%
Indicator 2S.4	90.9%
Indicator 2S.5	72.7%
Indicator 2S.6	72.7%

Indicator 2S.7	45.5%
Indicator 2S.8	77.3%
<b>Gateway 3 Indicators:</b>	
Indicator 3.1	54.5%
Indicator 3.2	95.5%
Indicator 3.3	86.4%
Indicator 3.4	77.3%
Indicator 3.5	54.5%
<b>Average Alignment Across All Indicators</b>	66.7%

**Table 4:** Scoring Rubric - Comprehensiveness

	Comprehensiveness	
	CVI (mean)	IRA (%)
<b>Indicators</b>		
Indicator 1.1	3.7	95.2%
Indicator 1.2	3.7	90.5%
<b>Gateway 1 Indicator Average</b>	3.7	92.9%
Indicator 2A.1	3.7	100%
Indicator 2A.2	3.4	90.5%
Indicator 2A.3	3.7	95.2%
Indicator 2A.4	3.7	90.5%
Indicator 2A.5	3.8	100%
Indicator 2A.6	3.8	95.2%
<b>Gateway 2 Adoption Indicator Average</b>	3.7	95.2%
Indicator 2L.1	3.7	100%

Indicator 2L.2	3.5	90.5%
Indicator 2L.3	3.8	100%
Indicator 2L.4	3.8	95.2%
Indicator 2L.5	3.8	95.2%
Indicator 2L.6	3.7	95.2%
Indicator 2L.7	3.9	95.2%
Indicator 2L.8	3.7	90.5%
Indicator 2L.9	3.9	95.2%
<b>Gateway 2 Launch Indicator Average</b>	3.8	95.2%
Indicator 2T.1	3.8	100%
Indicator 2T.2	3.6	95.2%
Indicator 2T.3	3.8	95.2%
Indicator 2T.4	3.8	100%
Indicator 2T.5	3.8	100%
Indicator 2T.6	3.9	100%
Indicator 2T.7	3.7	95.2%
<b>Gateway 2 Ongoing Professional Learning for Teachers Indicator Average</b>	3.8	97.9%
Indicator 2S.1	3.7	95.2%
Indicator 2S.2	3.5	90.5%
Indicator 2S.3	3.8	100%
Indicator 2S.4	3.8	95.2%
Indicator 2S.5	3.6	85.7%
Indicator 2S.6	3.8	95.2%
Indicator 2S.7	3.7	90.5%
Indicator 2S.8	3.8	95.2%
<b>Gateway 2 System Design and Leadership Support Indicator Average</b>	3.7	93.4%

<b>Gateway 2 Indicator Average (across all Gateway 2 constructs)</b>	3.8	95.4%
Indicator 3.1	3.8	100%
Indicator 3.2	3.7	95.2%
Indicator 3.3	3.9	100%
Indicator 3.4	3.8	95.2%
Indicator 3.5	3.7	90.5%
<b>Gateway 3 Indicator Average</b>	3.8	96.2%

**Table 5:** Scoring Rubric - Weighting

	Weighting			IRA (%)
	Should be weighted less than other Indicators (%)	Should be weighted more than other Indicators (%)	Should be weighted same as other Indicators (%)	
<b>Indicators</b>				
Indicator 1.1	33.3%	33.3%	33.3%	71.4%
Indicator 1.2	0%	75%	25%	76.2%
<b>Gateway 1 Indicator Average</b>	16.7%	54.15	29.2%	73.8%
Indicator 2A.1	66.7%	0%	33.3%	81%
Indicator 2A.2	25%	25%	50%	61.9%
Indicator 2A.3	22.2%	0%	77.8%	57.1%
Indicator 2A.4	20%	60%	20%	76.2%
Indicator 2A.5	20%	40%	40%	76.2%
Indicator 2A.6	0%	83.3%	16.7%	71.4%
<b>Gateway 2 Adoption Indicator Average</b>	25.7%	34.7%	39.6%	70.6%
Indicator 2L.1	75%	0%	25%	81%
Indicator 2L.2	30%	40%	30%	52.4%
Indicator 2L.3	25%	50%	25%	81%
Indicator 2L.4	0%	60%	40%	76.2%

Indicator 2L.5	0%	20%	80%	76.2%
Indicator 2L.6	0%	50%	50%	71.4%
Indicator 2L.7	0%	50%	50%	81%
Indicator 2L.8	71.4%	0%	28.6%	66.7%
Indicator 2L.9	0	37.5%	62.5%	61.9%
<b>Gateway 2 Launch Indicator Average</b>	22.4%	34.2%	43.5%	72%
Indicator 2T.1	50%	0%	50%	90.5%
Indicator 2T.2	33.3%	33.3%	33.3%	57.1%
Indicator 2T.3	0%	75%	25%	81%
Indicator 2T.4	0%	42.9%	57.1%	66.7%
Indicator 2T.5	0%	25%	75%	81%
Indicator 2T.6	0%	66.7%	33.3%	85.7%
Indicator 2T.7	0%	44.4%	55.6%	57.1%
<b>Gateway 2 Ongoing Professional Learning for Teachers Indicator Average</b>	11.9%	41%	47%	74.2%
Indicator 2S.1	4.8%	0%	95.2%	0%
Indicator 2S.2	9.5%	19%	71.4%	0%
Indicator 2S.3	0%	14.3%	85.7%	0%
Indicator 2S.4	0%	14.3%	85.7%	0%
Indicator 2S.5	4.8%	14.3%	81%	0%
Indicator 2S.6	0%	28.6%	71.4%	0%
Indicator 2S.7	0%	23.8%	76.2%	0%
Indicator 2S.8	4.8%	14.3%	81%	0%
<b>Gateway 2 System Design and Leadership Support Indicator Average</b>	3%	16.1%	81%	0%
<b>Gateway 2 Indicator Average (across all Gateway 2 constructs)</b>	15.8%	31.5%	52.8%	54.2%
Indicator 3.1	20%	20%	60%	76.2%
Indicator 3.2	0%	42.9%	57.1%	66.7%



Indicator 3.3	0%	25%	75%	81%
Indicator 3.4	0%	25%	75%	81%
Indicator 3.5	50%	16.7%	33.3%	71.4%
<b>Gateway 3 Indicator Average</b>	14%	25.9%	60.1%	75.3%